



Highlights

- Mühelose Untersuchung von Daten aus unterschiedlichen Perspektiven.
 - Schnelle Diagnose von Problemen mit fehlenden Daten.
 - Ersetzen fehlender Werte durch Schätzwerte.
 - Anzeige fehlender Datentypen und aller Extremwerte.
 - Entfernen nicht erkennbarer Verzerrungen.
-

IBM SPSS Missing Values

Bessere Modelle durch Vervollständigung fehlender Werte

Umfrage- und Marktforscher, Sozialwissenschaftler, Fachleute für Data-Mining und viele andere professionelle Anwender verlassen sich auf IBM SPSS Missing Values, um ihre Forschungsdaten zu prüfen.

Wenn Sie fehlende Daten ignorieren oder ausschließen, laufen Sie Gefahr, verzerrte oder nicht aussagekräftige Ergebnisse zu erhalten. Mit IBM SPSS Missing Values können Sie die fehlenden Werte imputieren und aussagekräftigere Schlüsse ziehen. IBM SPSS Missing Values ist ein wichtiges Tool für alle jene, die auf die Datengültigkeit Wert legen. Sie können die Daten ohne großen Aufwand auf Muster fehlender Daten untersuchen. Anschließend können Sie Schätzungen anhand einer Übersichtstabelle durchführen und fehlende Werte mit statistischen Algorithmen imputieren.

Sie können zum Beispiel in Umfragen die Fragen verbessern, die Sie anhand der beobachteten Muster fehlender Werte bereits als potenziell verwirrend eingestuft haben. Mithilfe der Tabelle mit den Prozentsätzen der nicht übereinstimmenden Muster können Sie sogar bestimmen, ob fehlende Werte für eine Variable in Zusammenhang mit fehlenden Werten einer anderen Variable stehen.

Möglicherweise finden Sie heraus, dass Befragte, die eine Frage zum Einkommen auch eine Frage zum Ausbildungsniveau umgehen. Nutzen Sie diese Informationen, um in Zukunft die Qualität der Umfragen zu verbessern.



Fehlende Daten schnell und einfach diagnostizieren

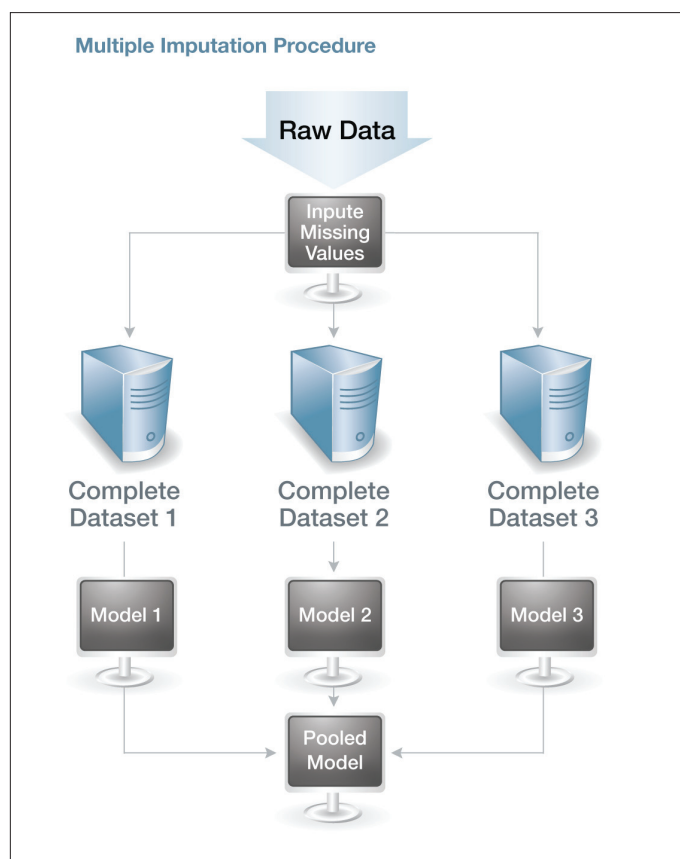
Mithilfe des Berichts zur Gesamtübersicht fehlender Werte können Sie ein schwerwiegendes Problem mit fehlenden Daten schnell diagnostizieren. Der Bericht über Muster fehlender Werte liefert Ihnen anschließend eine fallweise Übersicht über die Daten. Diese ist eine Kurzdarstellung der einzelnen Typen fehlender Daten und alle Extremwerte für die einzelnen Fälle.

Verwenden Sie den flexibel einsetzbaren T-Test bei separater Varianz (Separate-Variance T-Test) und Kreuztabellen für Tabellen mit kategorialen Variablen, um festzustellen, ob zwischen Befragten, die auf die Umfrage geantwortet haben, und jenen, die nicht geantwortet haben, signifikante Unterschiede bestehen. Anhand dieser Berichte können Sie entscheiden, ob fehlende Daten in der Analyse zu Problemen führen.

In den Berichten über tabellarische Muster erhalten Sie eine Zusammenfassung der Muster fehlender Daten und die wichtigsten Variablenätze, welche die Muster enthalten.

Fehlende Datenwerte durch Mehrfachimputation ersetzen

In IBM SPSS Missing Values werden Sie mit Verfahren zur Mehrfachimputation dabei unterstützt, Muster fehlender Daten im Dataset zu erkennen und fehlende Werte durch plausible Schätzwerte zu ersetzen. Das Produkt bietet einen vollautomatischen Imputationsmodus, der die am besten geeignete Imputationsmethode auf der Basis von Merkmalen der Daten auswählt. Sie haben zudem die Möglichkeit, das Imputationsmodell anzupassen.



Es werden einige vollständige Datasets generiert (normalerweise drei bis fünf), die jeweils einen anderen Satz von Ersetzungswerten aufweisen. Anschließend können Sie die einzelnen Datasets mithilfe der normalen Verfahren modellieren, zum Beispiel mit der linearen Regression, um Parameterschätzungen für die einzelnen Datasets zu erzeugen. Danach erhalten Sie endgültige Parameterschätzungen. Dazu zählen das Zusammenfassen (Pooling) der einzelnen Sätze von Parameterschätzungen aus Schritt zwei und das Berechnen der inferenziellen Statistik, die Variation innerhalb und zwischen Imputationen berücksichtigt.

Die Analyse der einzelnen Datasets und das Zusammenfassen der Ergebnisse werden über die Auswahl vorhandener IBM SPSS Statistics-Verfahren wie der Regression unterstützt. Wenn Sie Datasets mit imputierten Werten verarbeiten, liefern vorhandene Verfahren automatisch in Gruppen zusammengefasste („gepoolte“) Parameterschätzungen.

Zuverlässigere Schlüsse ziehen

Ersetzen Sie fehlende Werte durch Schätzwerte und erhöhen Sie die Wahrscheinlichkeit, statistisch signifikante Ergebnisse zu erhalten. Entfernen Sie nicht erkennbare Verzerrungen aus Ihren Daten, indem Sie fehlende Werte durch Schätzwerte ersetzen, sodass Sie alle Gruppen in die Analyse einbeziehen können – selbst Gruppen mit niedriger Rücklaufquote.

IBM SPSS Missing Values gehört zur IBM SPSS Statistics-Software-Suite. Die Produkte sind sowohl einzeln als auch in drei Editionen erhältlich: IBM SPSS Statistics Standard, IBM SPSS Statistics Professional und IBM SPSS Statistics Premium. Diese Editionen fassen wichtige Funktionalität zusammen. So können Sie effizient sicherstellen, dass Ihrem gesamten Team oder Ihrer Abteilung alle Funktionen zu Verfügung stehen, die sie benötigen, um die Analysen zu erstellen, die zum Erfolg Ihres Unternehmens beitragen.

IBM SPSS Missing Values ist als reine Client-Software verfügbar. Für eine höhere Leistung und Skalierbarkeit ist darüber hinaus eine serverbasierte Version verfügbar.

Wertzuwachs durch Zusammenarbeit

Damit Sie Assets effizient gemeinsam nutzen und wiederverwenden können, müssen diese so geschützt werden, dass interne und externe Konformitätsanforderungen erfüllt werden. Außerdem müssen die Ergebnisse so veröffentlicht werden, dass mehr Geschäftsbenutzer die Ergebnisse anzeigen und nutzen können. Zu diesem Zweck können Sie IBM SPSS Statistics durch IBM SPSS Collaboration and Deployment Services ergänzen. Weitere Informationen zu diesen wertvollen Funktionen erhalten Sie unter: ibm.com/spss/cds

Funktionen

Analyse von Mustern

- Zeigen Sie mithilfe der Tabelle für Datenmuster fehlende Daten und Extremfälle für alle Fälle und für alle Variablen an:
 - Systemdefinierte fehlende Werte und drei Typen benutzerdefinierter fehlender Werte anzeigen
 - In auf- oder absteigender Reihenfolge sortieren
 - Tatsächliche Werte für angegebene Variablen anzeigen
- Zeigen Sie mithilfe der Tabelle für Muster fehlender Daten alle Fälle an, die mindestens einen fehlenden Wert aufweisen:
 - Ähnliche Muster fehlender Daten gruppieren
 - Nach Mustern fehlender Daten und Variablen sortieren
 - Tatsächliche Werte für angegebene Variablen anzeigen
- Bestimmen Sie mithilfe der Tabelle für T-Test bei separater Varianz (Separate-Variance T-Test) Unterschiede zwischen fehlenden und nicht fehlenden Gruppen für eine in Beziehung stehende Variable:
 - T-Test, Freiheitsgrade, Mittelwert, Wert und Anzahl
- Zeigen Sie Unterschiede zwischen vorhandenen und fehlenden Daten für kategoriale Variablen mithilfe der Tabelle der Verteilung kategorialer Variablen an:
 - Kreuztabellen erstellen, die Produktdaten und fehlende Merkmalsausprägungen für jede Kategorie einer einzelnen Variablen in Bezug auf andere Variablen anzeigen

- Beurteilen Sie mithilfe der Tabelle für die Prozentsätze der nicht übereinstimmenden Muster, wie viele fehlende Daten für eine Variable mit den fehlenden Daten einer anderen Variable in Zusammenhang steht:
 - Matrizen nach Muster fehlender Werte oder Variablen sortieren
- Ermitteln Sie mithilfe der Tabelle für Muster in Tabellen alle eindeutigen Muster. Diese Tabelle fasst die einzelnen Muster fehlender Daten zusammen und zeigt für diese die Anzahl, Mittelwerte und Häufigkeiten für die einzelnen Variablen an:
 - Anzahl und Durchschnittswerte für die einzelnen Muster fehlender Werte mithilfe der Zusammenfassung der Tabelle für Muster fehlender Werte anzeigen

Statistik

- Univariat: Anzahl, Mittelwert, Standardabweichung und Standardfehler des Mittelwerts für alle Fälle berechnen außer für diejenigen, die fehlende Werte, Anzahl, Prozentsatz fehlender Werte und Extremwert für alle Variablen enthalten
- Listenweise: Mittelwert, Kovarianzmatrix und Korrelationsmatrix für alle quantitativen Variablen für Fälle berechnen und dabei fehlende Werte ausschließen
- Paarweise: Häufigkeit, Mittelwert, Varianz, Kovarianzmatrix und Korrelationsmatrix berechnen

Mehrfachimputation

- Geben Sie an, welche Variablen imputiert werden sollen, welche Bedingungen für die imputierten Werte gelten sollen, z. B. Mindest- und Höchstwert, und welche Variablen als Einflussvariablen zu verwenden sind, wenn fehlende Werte anderer Variablen imputiert werden.
- Imputieren Sie Werte für kategoriale und kontinuierliche Variablen: Die logistische Regression wird für kategoriale Variablen und die lineare Regression für kontinuierliche Variablen verwendet. Die Verwendung des am nächsten liegenden entfernten beobachteten Werts („Predictive Mean Matching“) ist eine Option für kontinuierliche Ergebnisse und stellt sicher, dass die imputierten Werte angemessen sind (sich im Bereich der Originaldaten befinden).
- Anhand der Mustererkennung für fehlende Daten kann das zu verwendende Imputationsverfahren bestimmt werden.
- Es stehen drei Imputationsverfahren zur Auswahl:
 - Monoton: effizientes Verfahren für Daten, die ein monotones Muster fehlender Werte aufweisen
 - Vollständig bedingte Spezifikation (FCS): iterative Markov-Ketten-Monte-Carlo-Verfahren (MCMC = Markov Chain Monte Carlo), das sich eignet, wenn die Daten willkürliche (monotone oder nicht monotone) Muster für fehlende Werte aufweisen
 - Automatisch: Durchsuchen der Daten, um das beste Imputationsverfahren (monoton oder FCS) zu bestimmen
- Sie können Folgendes angeben:
 - Anzahl der Imputationen
 - Bereich der imputierten Werte
 - Ob beim Imputieren Wechselwirkungen verwendet werden
 - Optional, ob die Imputation für Variablen inaktiviert werden soll, die einen höheren Prozentsatz fehlender Werte aufweisen
 - Toleranzstufen zum Prüfen auf Singularität
- Geben Sie eine Variable an, die Gewichte für die Analyse (Regression) enthält: Bei dieser Vorgehensweise werden Analysegewichte in Regressions- und Klassifikationsmodelle integriert, die zum Imputieren fehlender Werte verwendet werden. Analysegewichte werden darüber hinaus in Zusammenfassungen imputierter Werte verwendet (z. B. Mittelwert, Standardabweichung und Standardfehler).
- Zeigen Sie eine Zusammenfassung für das Fehlen von Daten an sowie eine Zusammenfassung der Imputation und das Imputationsmodell für die einzelnen Variablen, deren Werte imputiert werden. Sie können eine nach Variablen strukturierte Analyse fehlender Werte sowie tabellarische Muster fehlender Werte oder eine deskriptive Statistik für imputierte Werte abrufen.
- Werten Sie das Fehlen von Werten für Fälle, Variablen und einzelne Datenwerte (Zellenwerte) grafisch aus.
- Fordern Sie eine IBM SPSS Statistics-Datendatei an, die imputierte Werte und/oder ein FCS-Iterationsprotokoll enthält.
- Datasets für Mehrfachimputation können mithilfe unterstützter Analyseverfahren analysiert werden, um endgültige (kombinierte) Parameterschätzungen zu erhalten, die die inhärente Ungewissheit in den verschiedenen Sätzen imputierter Werte berücksichtigen.

Analyse

- Unterstützte Analyseverfahren für Mehrfachimputation (Anmerkung: Sie müssen das geeignete Modul erworben haben, das das Verfahren enthält.)
- Deskriptive Verfahren: Häufigkeiten, deskriptive Statistik, Kreuztabellen, Korrelationen, nicht parametrische Korrelation, partielle Korrelation
- Vergleich von Mittelwerten: Mittelwert, T-Test, nicht parametrische Tests, einfaktorielle Varianzanalyse (einfaktorielle ANOVA), univariate Varianzanalyse (univariate ANOVA)
- Modelle: allgemeine lineare Modelle, verallgemeinerte lineare Modelle, lineare Regression, multinomiale logistische Regression, binäre logistische Regression, Diskriminanzanalyse, ordinale Regression, lineare gemischte Modelle
- Verfahren der Überlebensanalyse: Cox-Regression

Zusammenfassen (Pooling)

- Zusammenfassen (Pooling) der Ausgabe: Die Ausgabe wird mithilfe einer von insgesamt zwei Zusammenfassungsstufen zusammengefasst und erzeugt zusammengefasste Parameter
- Zusammenfassungsdiagnose:
 - Relative Zunahme der Varianz: Maß der imputationsübergreifenden relativen Variabilität der Parameterschätzung
 - Anteil der fehlenden Daten: relative anteilmäßige Zunahme der Varianz; Maß der Ungewissheit aufgrund fehlender Antworten
 - Relative Effizienz: Effizienz der Schätzung für M Imputationen im Vergleich zu derjenigen für eine unendliche Anzahl von Imputationen
- Abruf der Modell-PMML für zusammengefasste Parameterschätzung: lineare Regression, allgemeine lineare Modelle, multinomiale logistische Regression, binäre logistische Regression, Diskriminanzanalyse, Cox-Regression

Systemvoraussetzungen

Die Anforderungen variieren je nach Plattform. Einzelheiten finden Sie unter [ibm.com/spss/requirements](https://www.ibm.com/spss/requirements)

Informationen zu IBM Business Analytics

IBM Business Analytics-Software stellt den Entscheidern verlässliche Informationen zur Verfügung, die für fundierte Entscheidungen nötig sind. IBM bietet ein umfassendes, einheitliches Portfolio für Business Intelligence, vorausschauende und erweiterte Analyse, Financial Performance- und Strategiemangement, Governance, Risikomanagement und Compliance sowie Analyseanwendungen.

Mit IBM Software können Unternehmen Trends, Muster und Unregelmäßigkeiten erkennen, „Was wäre, wenn“-Szenarien vergleichen, mögliche Bedrohungen und Chancen vorhersagen, kritische Geschäftsrisiken erkennen und minimieren sowie Ressourcen planen, budgetieren und prognostizieren. Durch diese umfassenden Analysefunktionen sind unsere Kunden rund um den Globus in der Lage, ihre Geschäftsergebnisse besser zu verstehen, vor auszusehen und zu beeinflussen.

Weitere Informationen

Wenn Sie mehr erfahren möchten, besuchen Sie uns unter:

ibm.com/de/spss



IBM Deutschland GmbH
IBM-Allee 1
71139 Ehningen
ibm.com/de

IBM Österreich
Obere Donaustrasse 95
1020 Wien
ibm.com/at

IBM Schweiz
Vulkanstrasse 106
8010 Zürich
ibm.com/ch

Die IBM Homepage finden Sie unter:
ibm.com

IBM, das IBM Logo, ibm.com und SPSS sind eingetragene Marken der IBM Corporation in den USA und/oder anderen Ländern. Weitere Produkt- und Servicennamen können Marken von IBM oder anderen Herstellern sein. Eine aktuelle Liste der IBM Marken finden Sie auf der Webseite „Copyright and trademark information“ unter ibm.com/legal/copytrade.shtml

Der Inhalt dieses Dokuments ist nur zum Datum der Erstveröffentlichung des Dokuments aktuell und kann jederzeit ohne vorherige Ankündigung geändert werden. Die IBM Angebote können von Land zu Land unterschiedlich sein.

Vertragsbedingungen und Preise erhalten Sie bei den IBM Geschäftsstellen und/oder den IBM Business Partnern. Die Produktinformationen geben den derzeitigen Stand wieder. Gegenstand und Umfang der Leistungen bestimmen sich ausschließlich nach den jeweiligen Verträgen.

© Copyright IBM Corporation 2012



Bitte der Wiederverwertung zuführen